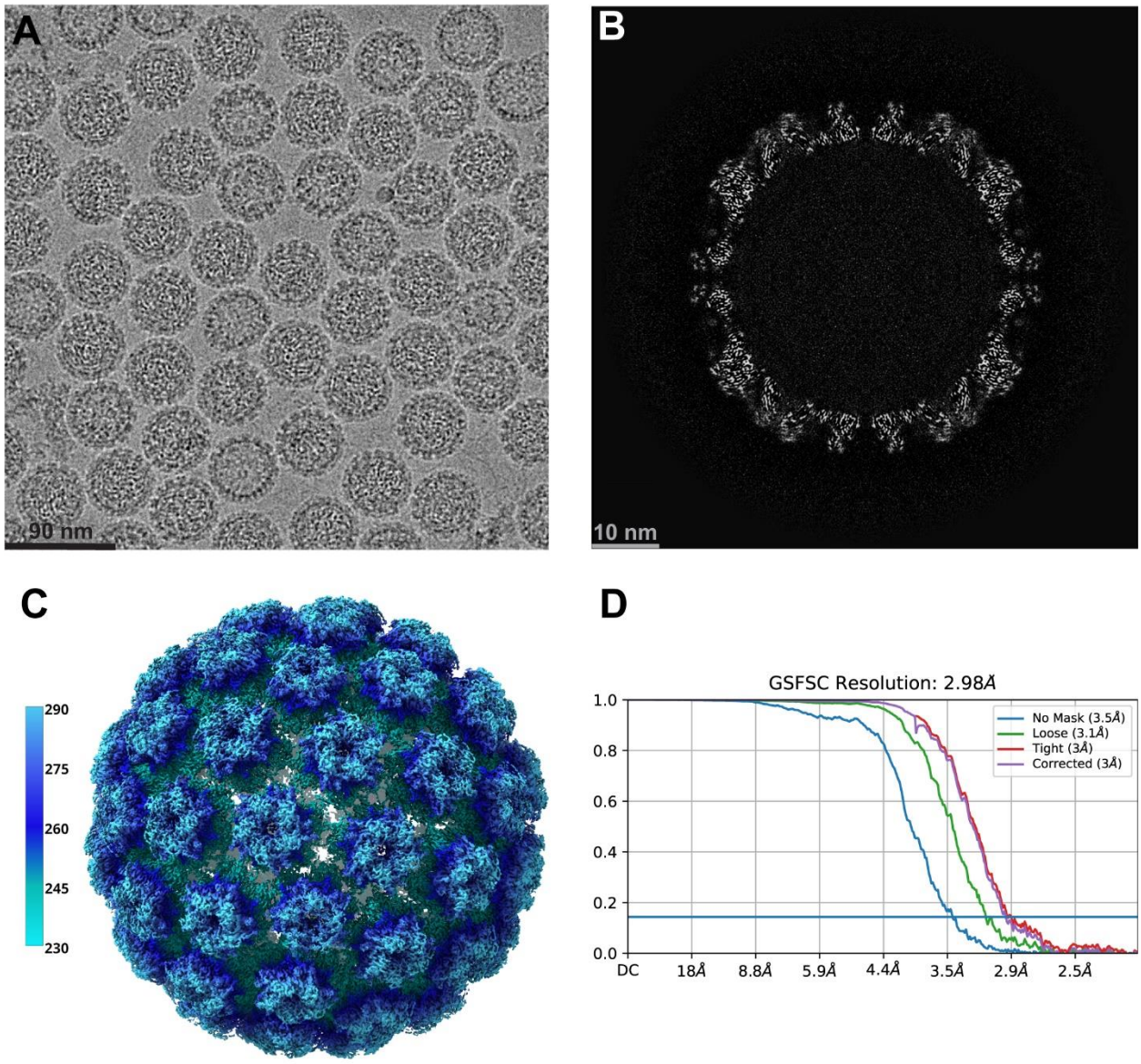
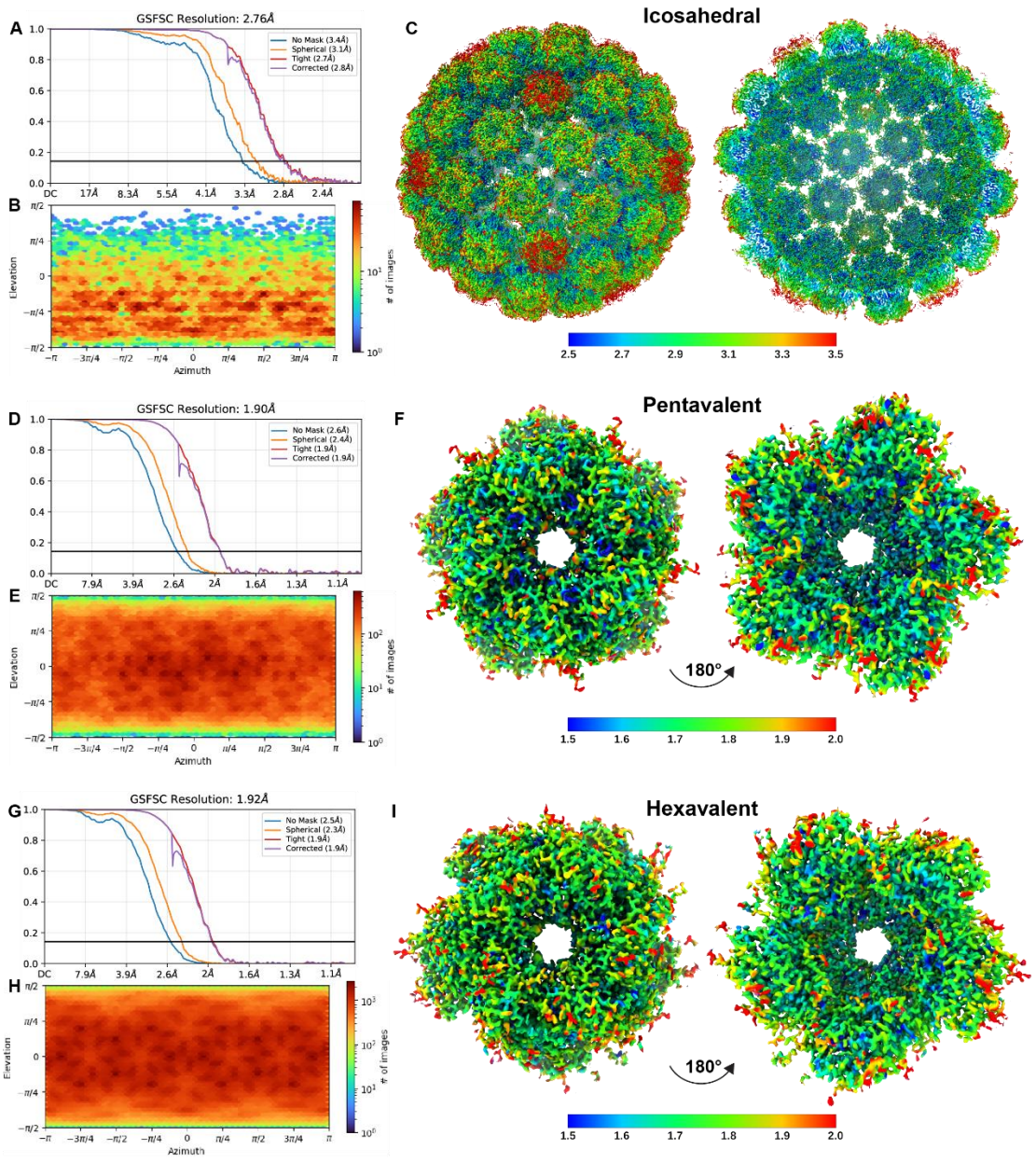
Supplementary Information for

Atomic resolution structure of human papillomavirus bound to heparin

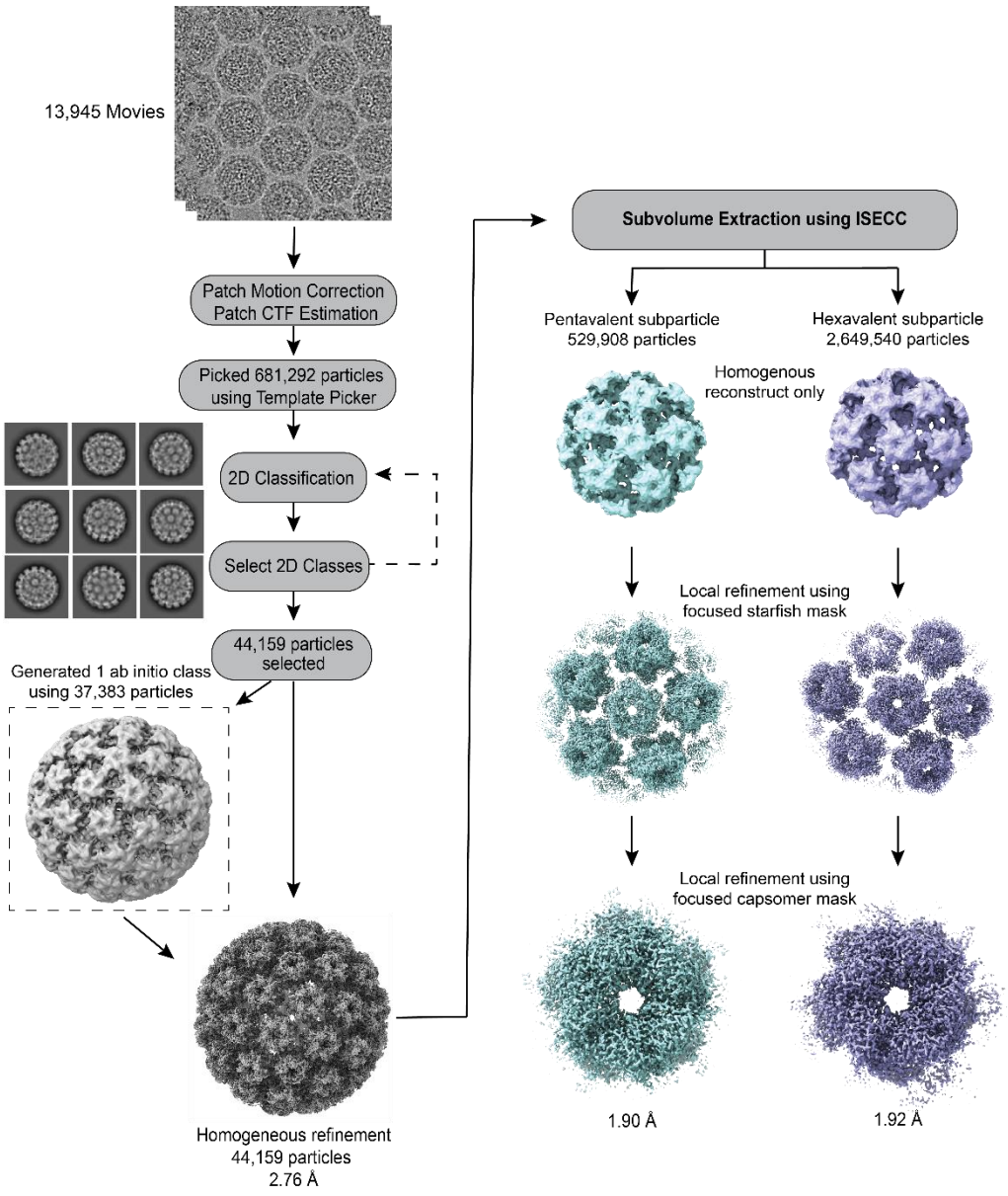
Table of Contents



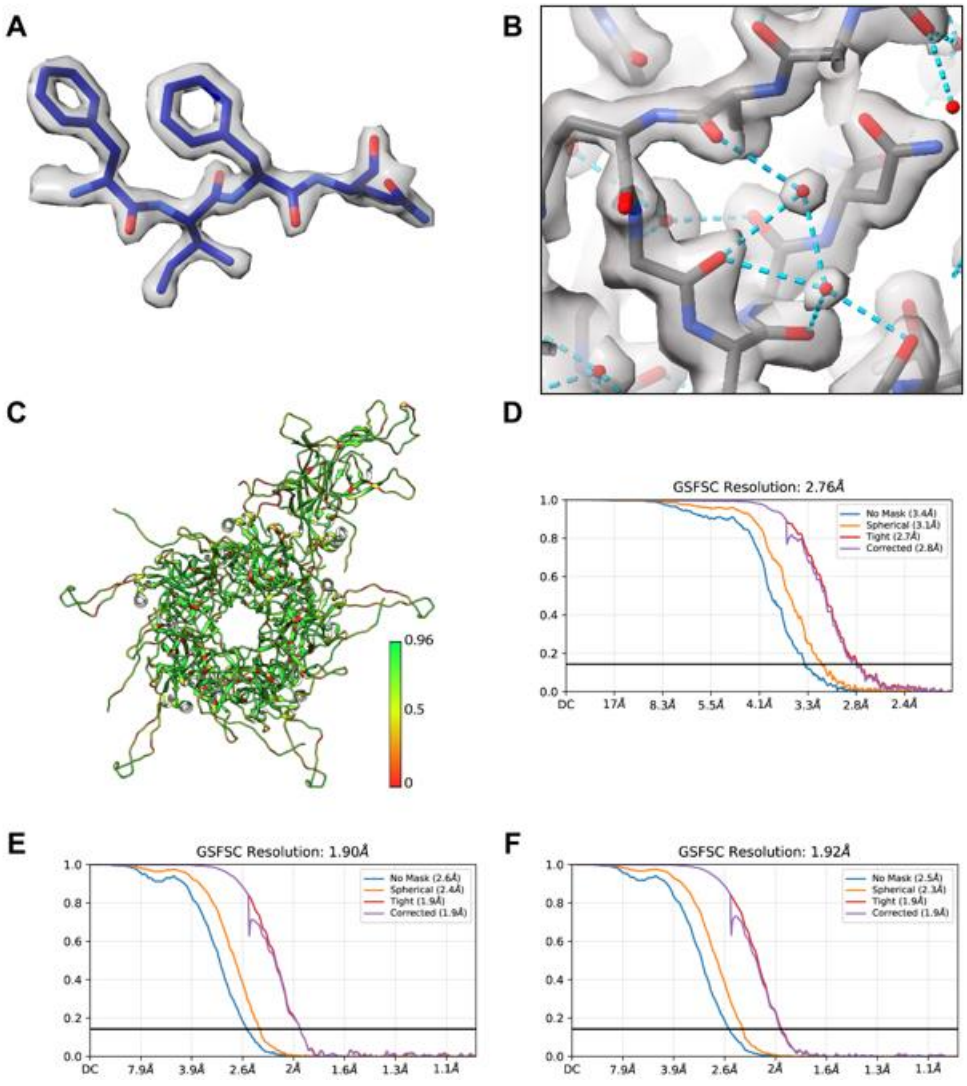
Supplementary Figure 1: 59kX data collection and cryoEM icosahedral map and refinement. (A) Representative micrograph of purified HPV16 quasivirus bound to heparin, scale bar 90nm. Data were collected from the same grids for both data collections, but at different magnifications. (B) Central slab from the icosahedral 3D refinement shows the quality of the map. Scale bar 10nm. (C) Icosahedrally averaged map at 2.98 Å resolution, colored radially in Å according to the color key (left). (D) FSC curves for the icosahedral refinement.



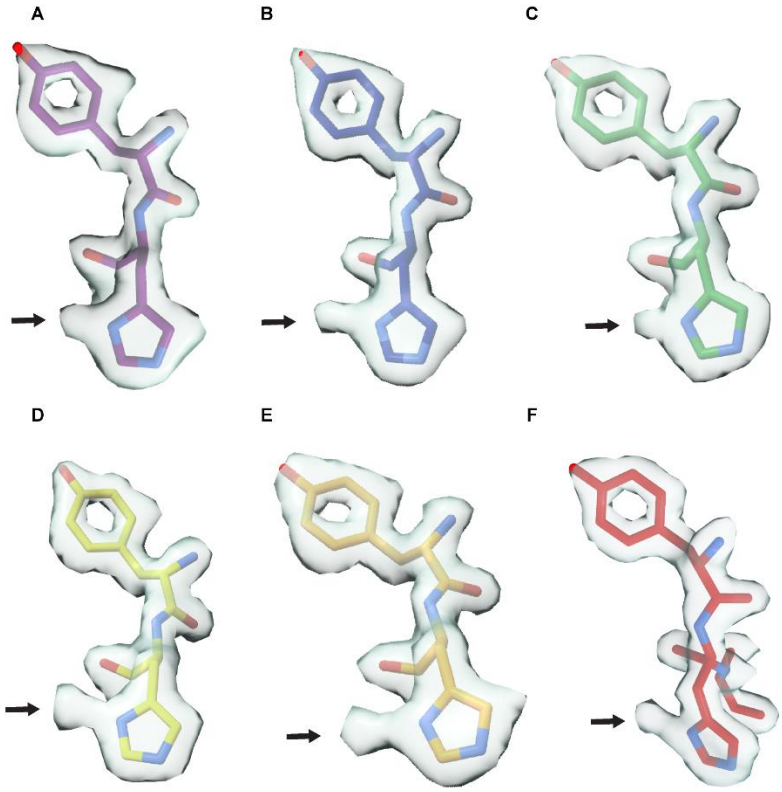
Supplementary Figure 2: Icosahedral and subparticle map local resolution and orientations. Gold-standard FSC curves for the 2.76 Å icosahedral (A), 1.90 Å pentavalent subparticle (D), and 1.92 Å hexavalent subparticle (G) maps. Viewing direction distribution graphs for the icosahedral (B), pentavalent subparticle (E), and hexavalent subparticle (H) maps. (C) HPV16 bound to heparin icosahedral map, with external and internal surfaces shown, was colored by resolution ranging from 2.5-3.5 Å, according to key below. (F) Pentavalent and (I) hexavalent maps with external and internal surfaces shown, colored by resolution ranging from 1.5-2.0 Å, according to key below each map.



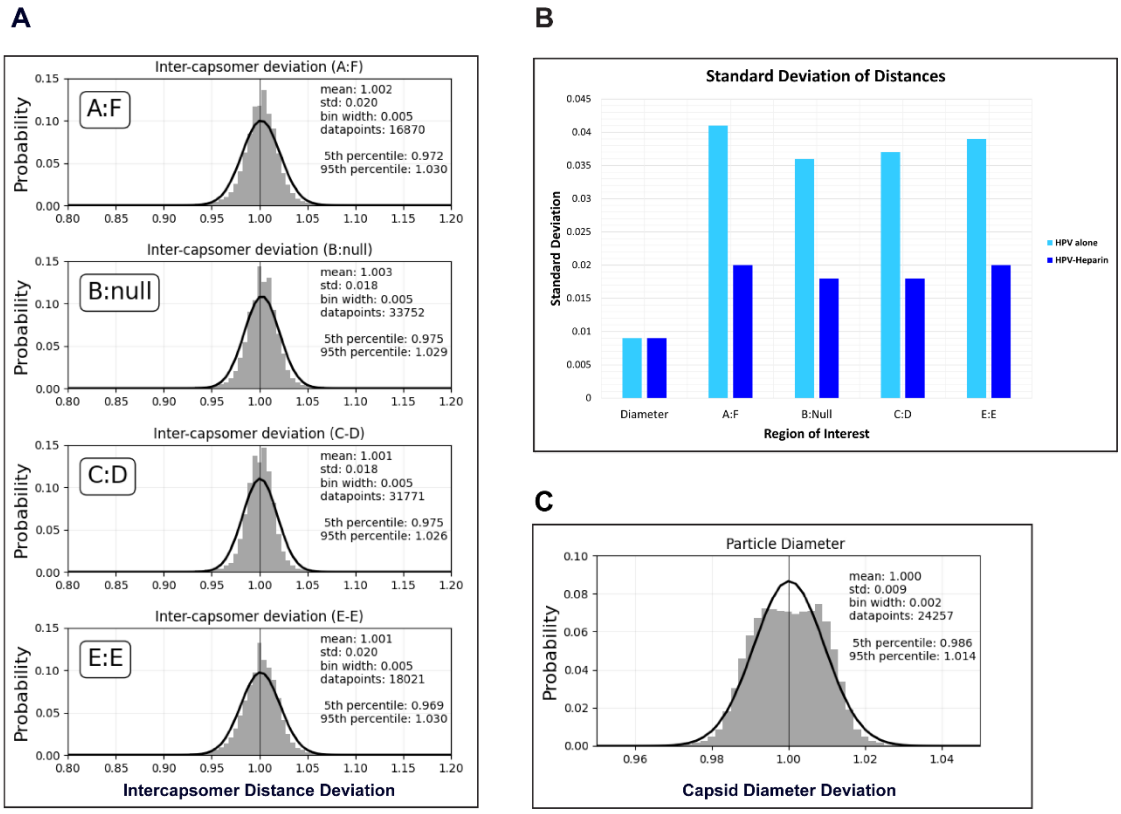
Supplementary Figure 3: CryoEM data processing for HPV16 at 0.54 Å pixel size. The refinement method shown in the flowchart includes processing of raw micrographs, 2D classification, homogenous refinement, subvolume extraction with ISECC, and 3D local refinement of the pentavalent and hexavalent subparticles.



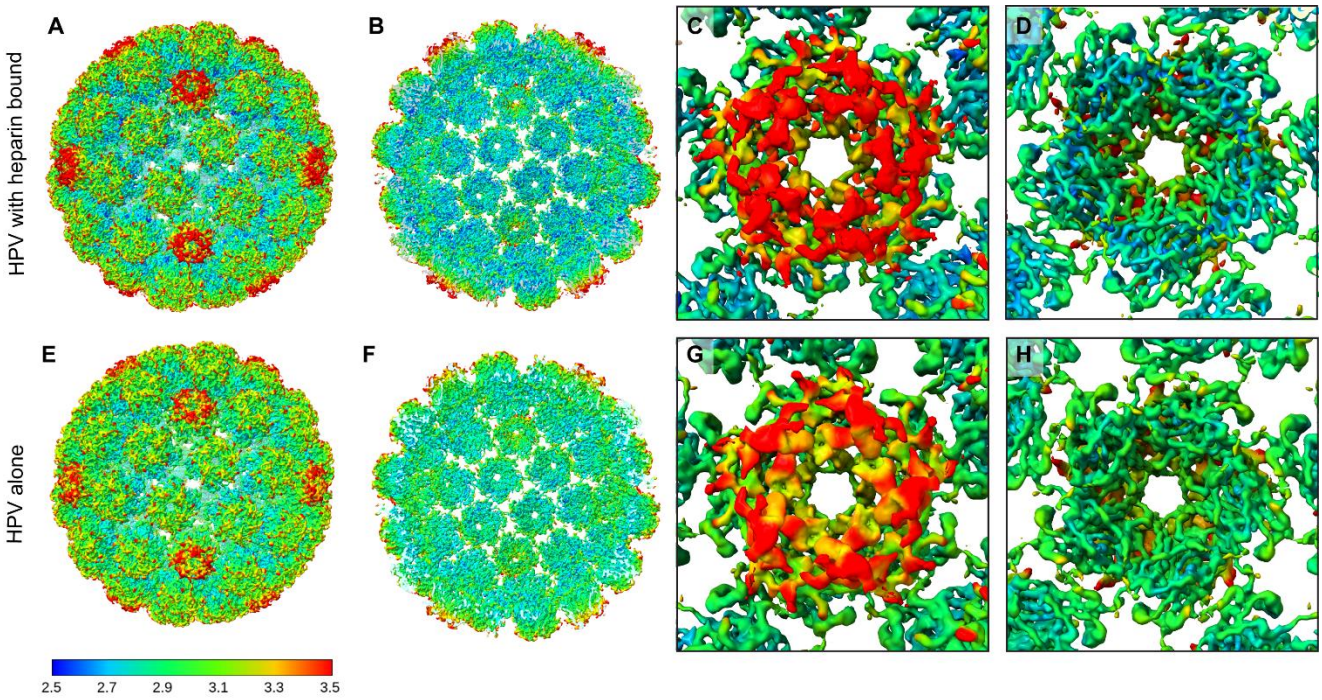
Supplementary Figure 4: HPV-heparin map and model quality analysis. (A) Example of map to model fit, with L1 residues 373-377 (blue) fit into the map (gray transparent). (B) Representative image of water molecules identified within the 1.9 Å map. Zoomed in view of L1 model (dark gray with colored heteroatoms) fit into the 1.9 Å map (transparent gray) to show water molecules. Hydrogen bonds (light blue) were identified between water molecules and the L1 model. (C) Model of the asymmetric unit colored by Q-score, with high Q-score in green and low Q-score in red. Overall Q-score of the model is 0.86, which correlates well considering the resolution of the map[33]. (D,E,F) FSC curves for the icosahedral, pentavalent subparticle, and hexavalent subparticle maps, respectively.



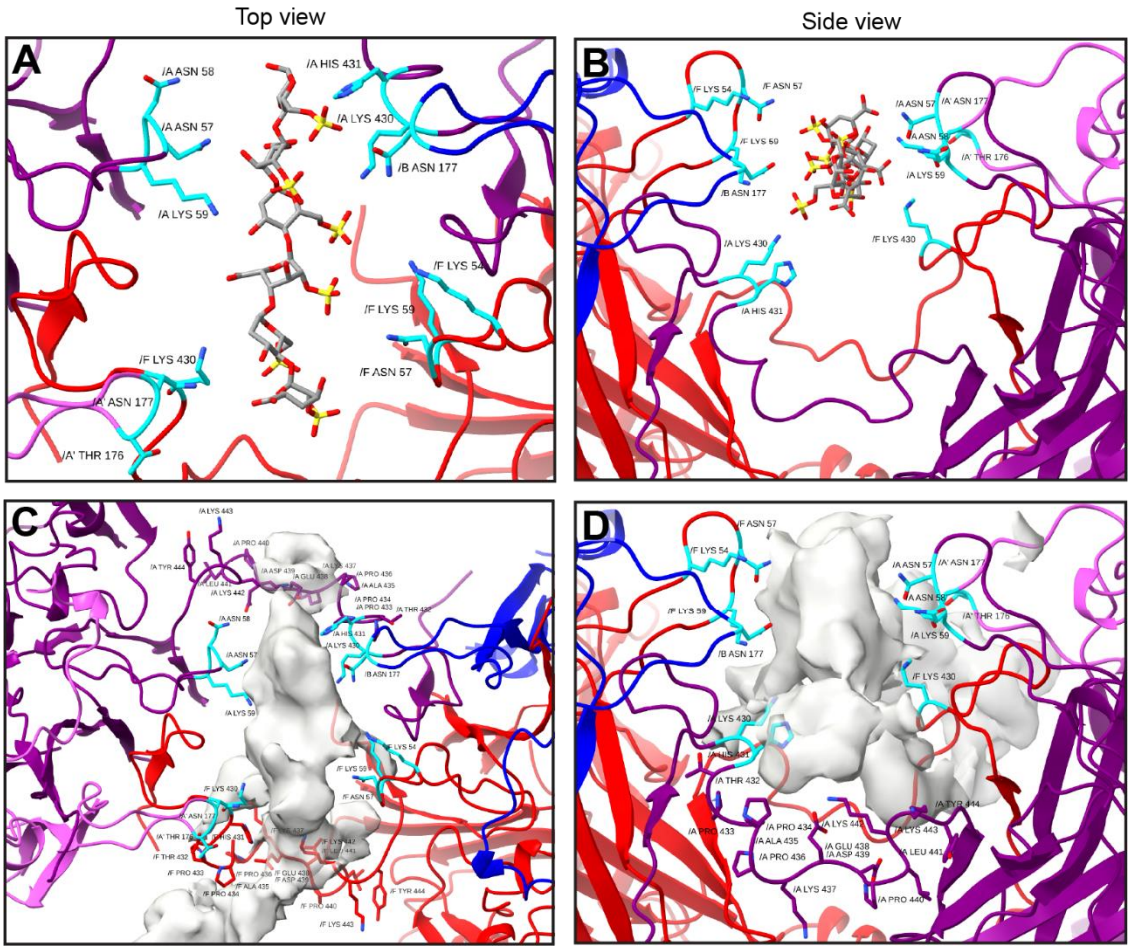
Supplementary Figure 5: Unfilled extra density at histidine 36 in all six chains of the asymmetric unit. (A-F) Tyrosine 35 and histidine 36 are shown for each of the six L1 chains A-F (purple, blue, green, yellow, orange, and red, respectively). Extra density (indicated with an arrow) is seen consistently on the histidine 36 ring.



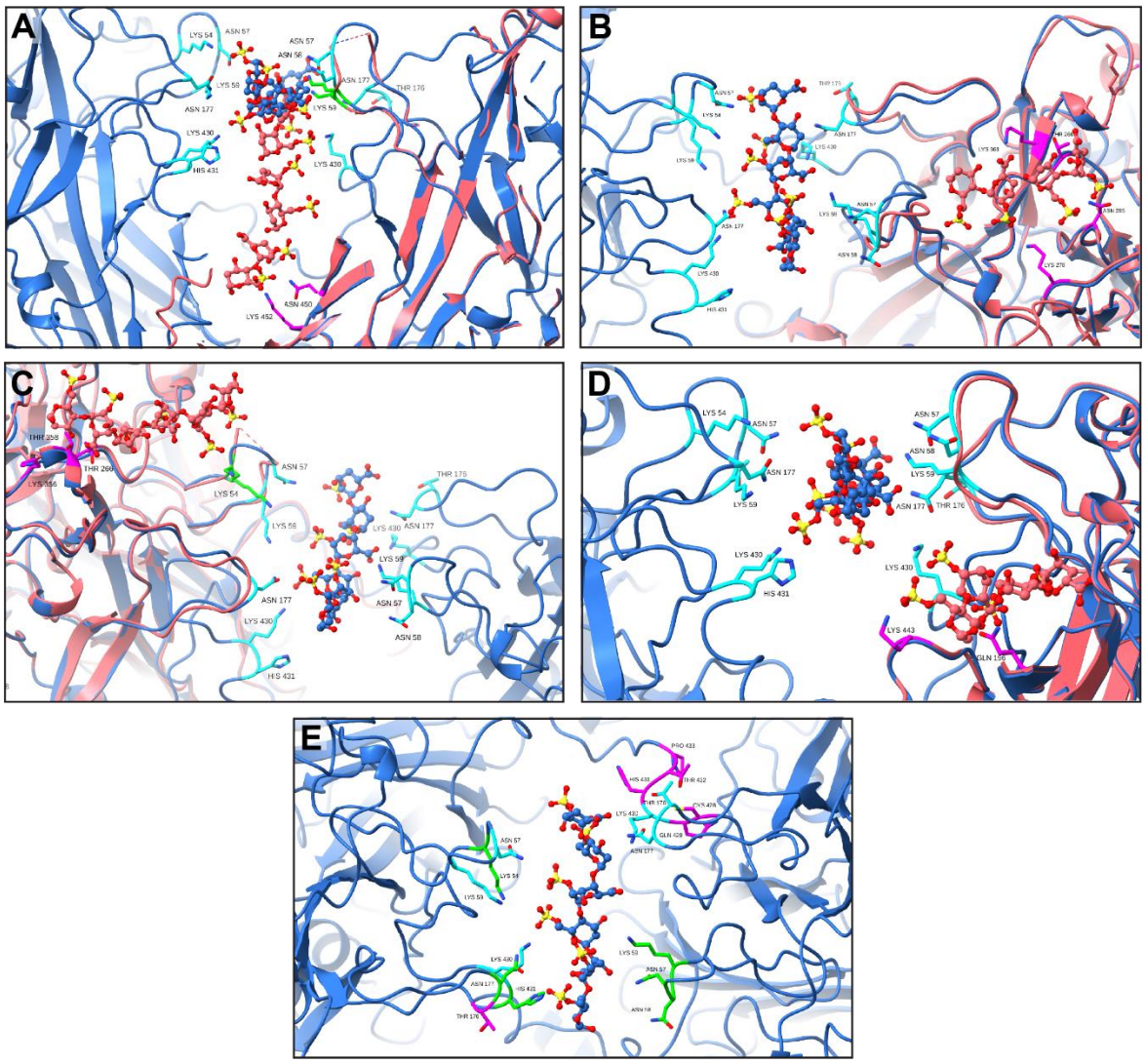
Supplementary Figure 6: ISECC flexibility analysis of arm connections and capsid diameter. (A) The distance between unique capsomers was measured and compared to icosahedrally imposed values, and the ratios of observed to expected distances are plotted on the x-axis[14]. Flexibility values do not follow the typical Gaussian distribution (black curve). A:F corresponds to the pentavalent and the closest hexavalent capsomer. B:null corresponds to the hexavalent capsomers surrounding the pentavalent capsomer. C:D corresponds to the capsomers surrounding the threefold axis. E:E corresponds to capsomers on either side of the twofold axis. (B) Bar chart of standard deviation of distances from the ISECC flexibility analysis. The standard deviation of distances for HPV bound to heparin (dark blue) were compared to the standard deviation of distances for HPV alone (light blue)[14]. Decreased standard deviation of distances in HPV bound to heparin indicates decreased flexibility. (C) The distance between capsomers on opposite sides of the capsid was measured and compared to icosahedrally imposed values, and the ratios of observed to expected distances are plotted on the x-axis[14]. Capsid diameters did not show a Gaussian distribution (black curve). Source data are provided as a Source Data file.



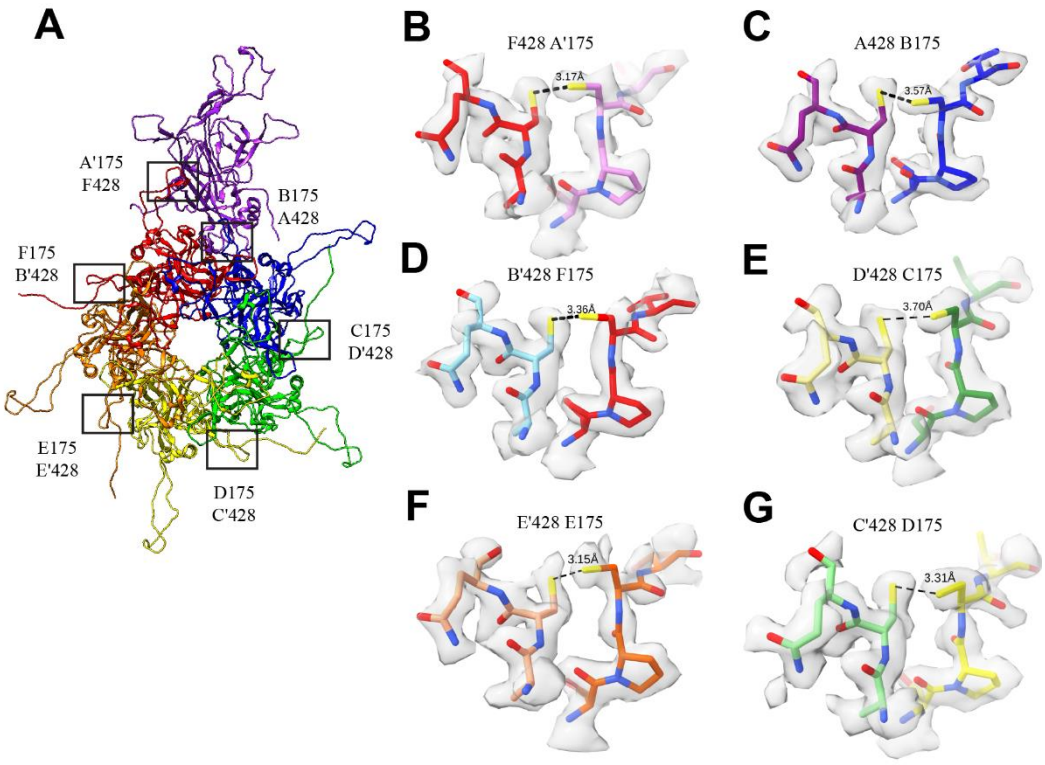
Supplementary Figure 7: Local resolution of icosahedral papillomavirus cryoEM maps. (A) External and (B) internal views of 2.98Å HPV heparin map shown with (C) external and (D) internal zoomed in views of a pentavalent capsomer. (E) External and (F) internal views of 3.1Å HPV map shown with (G) external and (H) internal zoomed in views of a pentavalent capsomer[14]. Both capsids were colored radially according to local resolution as shown by the scale bar on the bottom left.



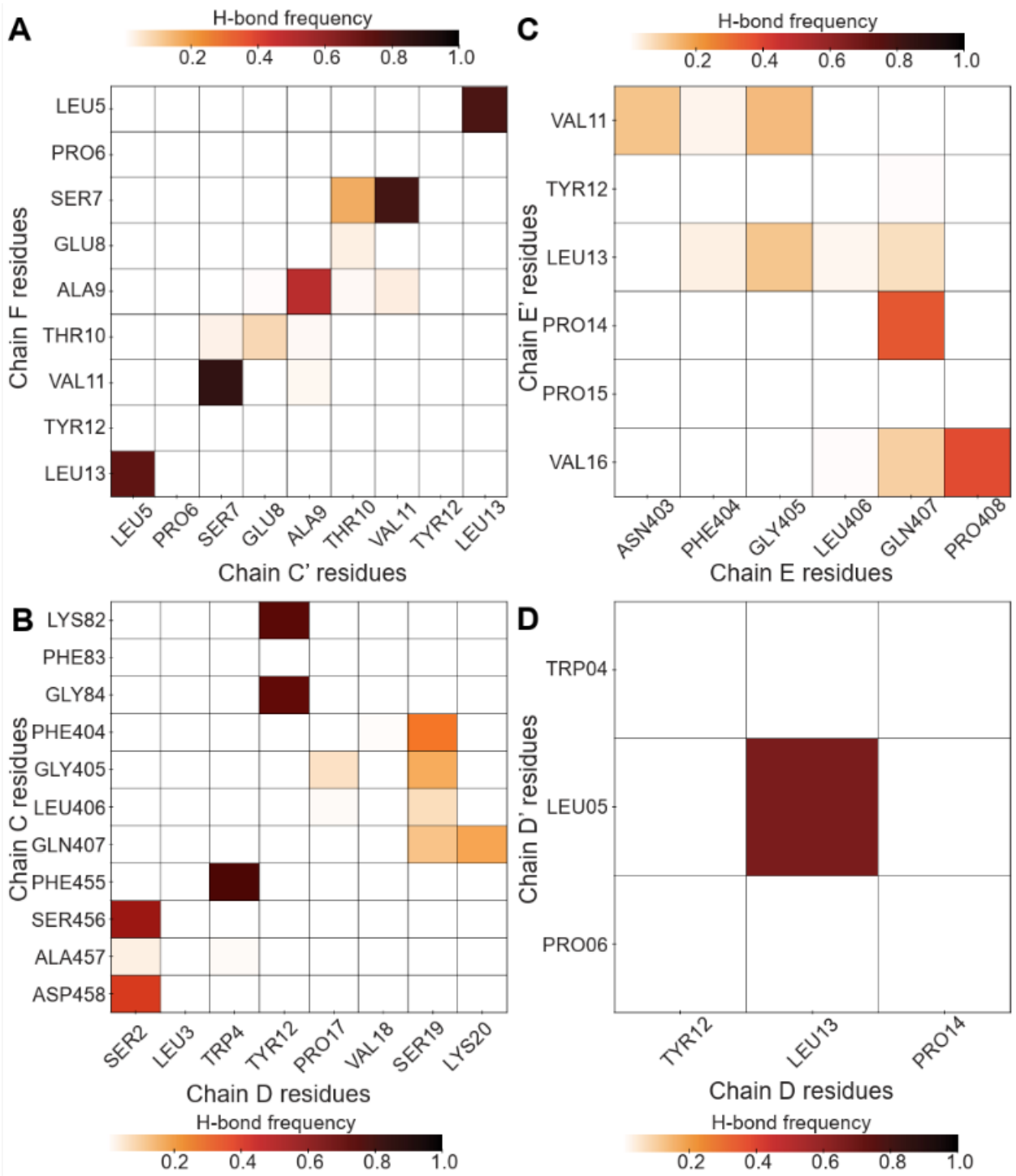
Supplementary Figure 8: Close up view of heparin binding site. (A,B,C,D) Four L1 proteins, Chain A (purple), Chain A' (light purple), Chain B (blue), and Chain F (red) are shown. Heparin binding site with residues that are likely interacting with heparin are shown in cyan. Amino acids interacting with heparin in each chain are: Chain F – Lys 54, Asn 57, Lys 59, Lys 430, Chain B – Asn 177, Chain A – Asn 57, Asn 58, Lys 59, Lys 430, His 431, Chain A' – Thr 176, Asn 177. (C,D) Heparin density at low contour (0.055) shown in transparent gray. Based on proximity to the heparin density, amino acids 431-444 in chains A and F may interact with heparin.



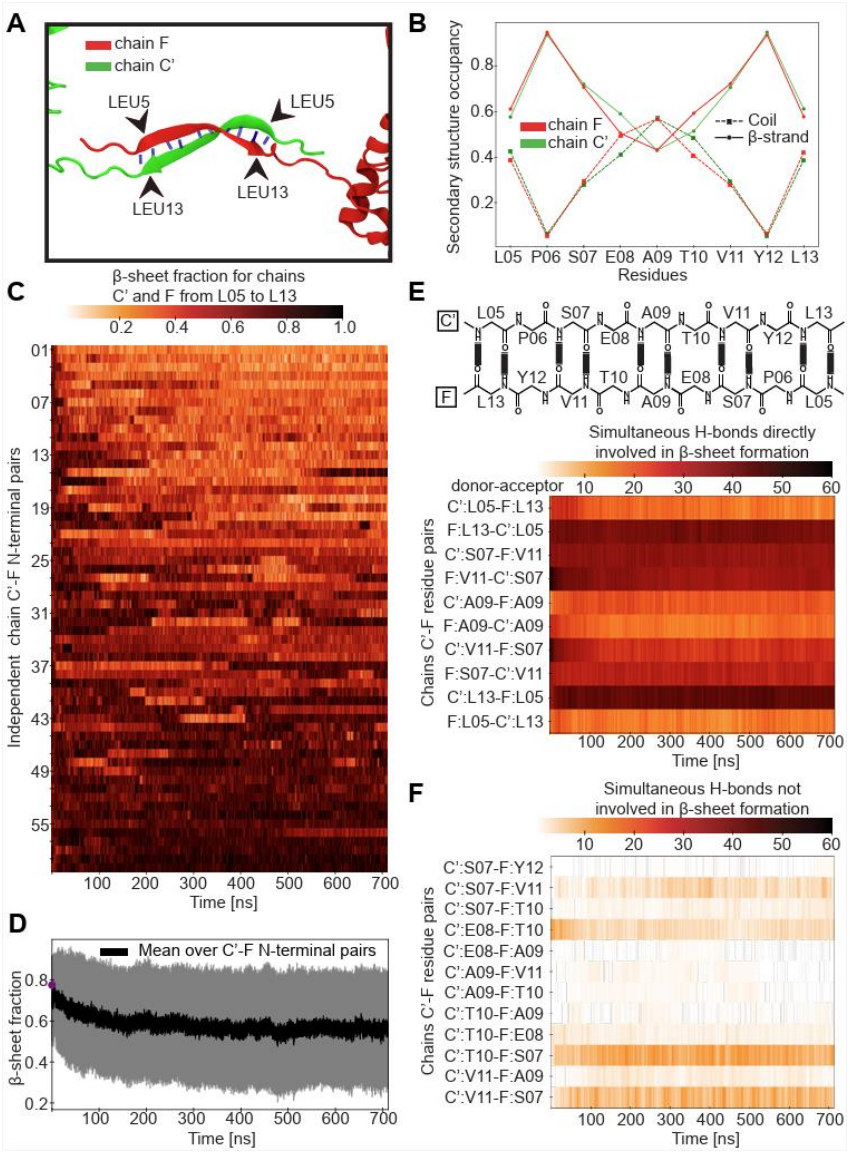
Supplementary Figure 9: Comparison of heparin binding sites and contact residues from previously solved structures. (A-D) Four HPV16 L1 chains (blue) and the pentamer modeled in Dasgupta et al. 2011 (PDB 5W1O) (red) are shown, oriented to view the gap between the hexavalent capsomer (left) and the pentavalent capsomer (right)[34]. PDB 5W1O is composed of pentamers of HPV16 L1, which were incubated with short 8-10 heparin subunits. The heparin built is shown in the respective colors, blue for the model presented here and red for PDB 5W1O, and the four heparin molecules modeled in PDB 5W1O are oriented towards the gap between the hexavalent/pentavalent capsomer, with a different heparin molecule and binding site for each panel A-D. The heparin binding sites were compared between this proposed binding site (cyan) and the four sites identified in Dasgupta et al. 2011 (magenta), with overlapping residues in neon green. Overall, there are two contact residues in common with the Dasgupta et al. 2011 paper (neon green, Panel A Lys 59, Panel C Lys 54). (E) Following the color scheme established in panels A-D, the heparin binding site from Guan et al. 2017 is shown in magenta, with the binding site proposed here in cyan and the overlapping residues in neon green[25]. There are six residues in common with the 4.3 Å resolution Guan result (neon green, Panel E, Lys 54, Asn 57, Asn 58, Lys 59, Asn 177, and His 431).



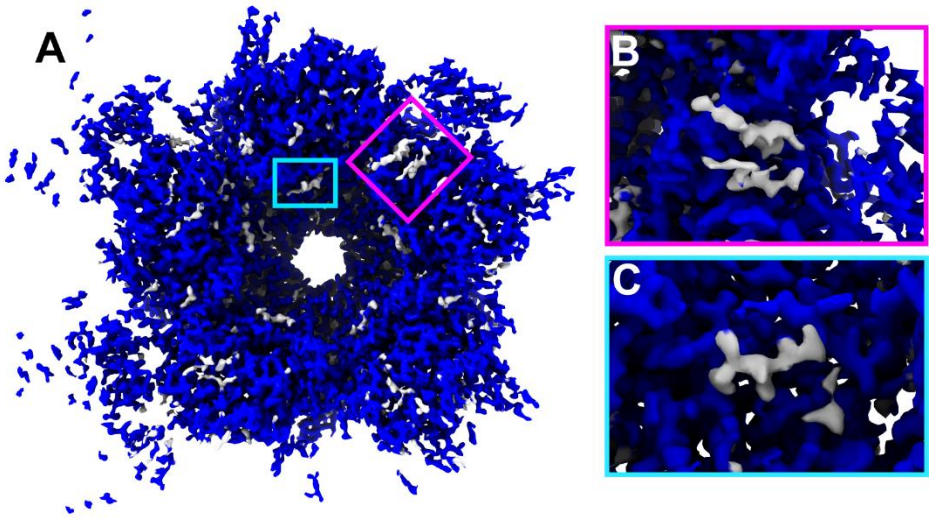
Supplementary Figure 10: CryoEM density corresponding to the disulfide bond between Cys 175 and Cys 428. (A) The asymmetric unit of HPV is shown with six L1 chains A-F colored purple, blue, green, yellow, orange, and red respectively to provide the six locations of paired cysteines. The location of the loop region containing Cys 175 for each L1 chain is indicated with a rectangle. (B-G) The cryoEM map density is shown at contour 0.779 (transparent gray) with L1 structure shown as stick at the location of each potential disulfide bond. The close up view shows each Cys 175 and the paired Cys 428 from the neighboring capsomer which provides the potential for disulfide bond formation[20]. The distance between sulfur atoms is shown for each in Å.



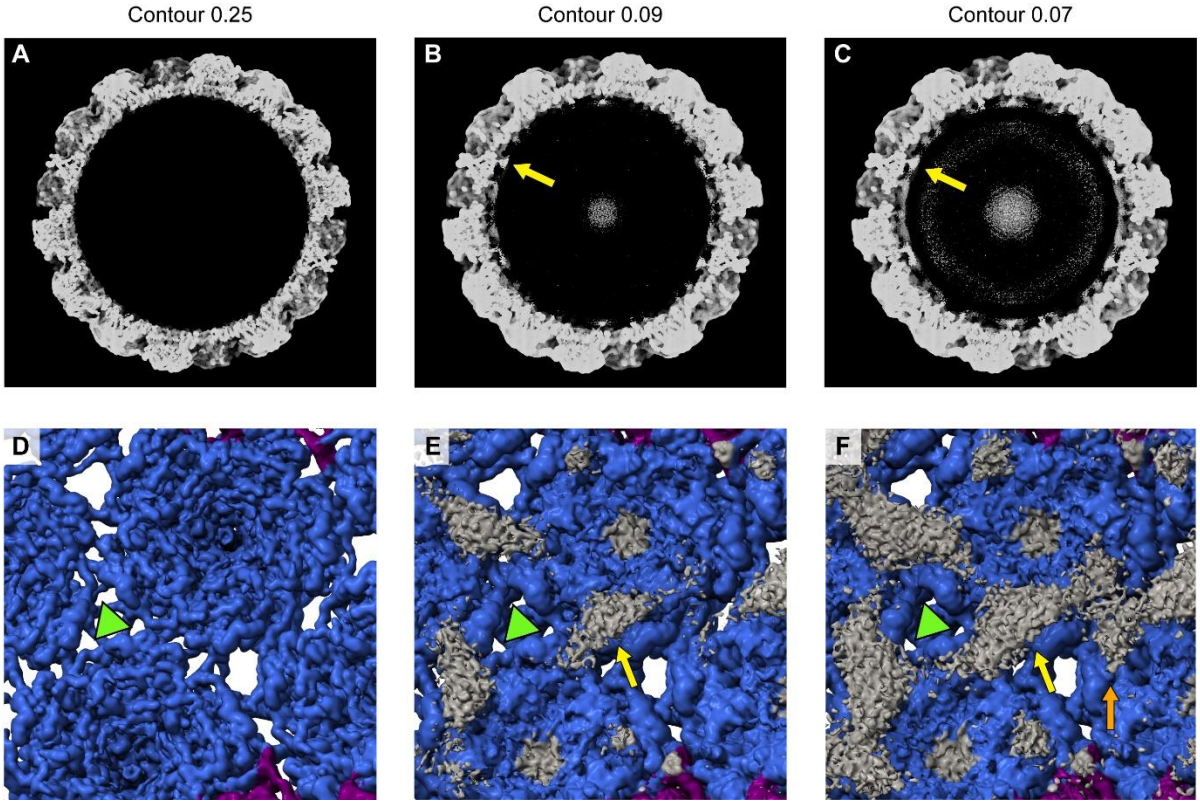
Supplementary Figure 11: Frequency of hydrogen bonds formed between N-terminal regions of L1 proteins. (A) chains C' and F, (B) chains C and D, (C) chains E and E', (D) chains D and D'. The frequency based on aggregate sampling is mathematically equivalent to the fractional average of the number of independent chain copies forming hydrogen bonds at any given time. Data corresponds to 42.6 μs of cumulative sampling aggregated over all 60 quasi-equivalent interactions. Source data are provided as a Source Data file.



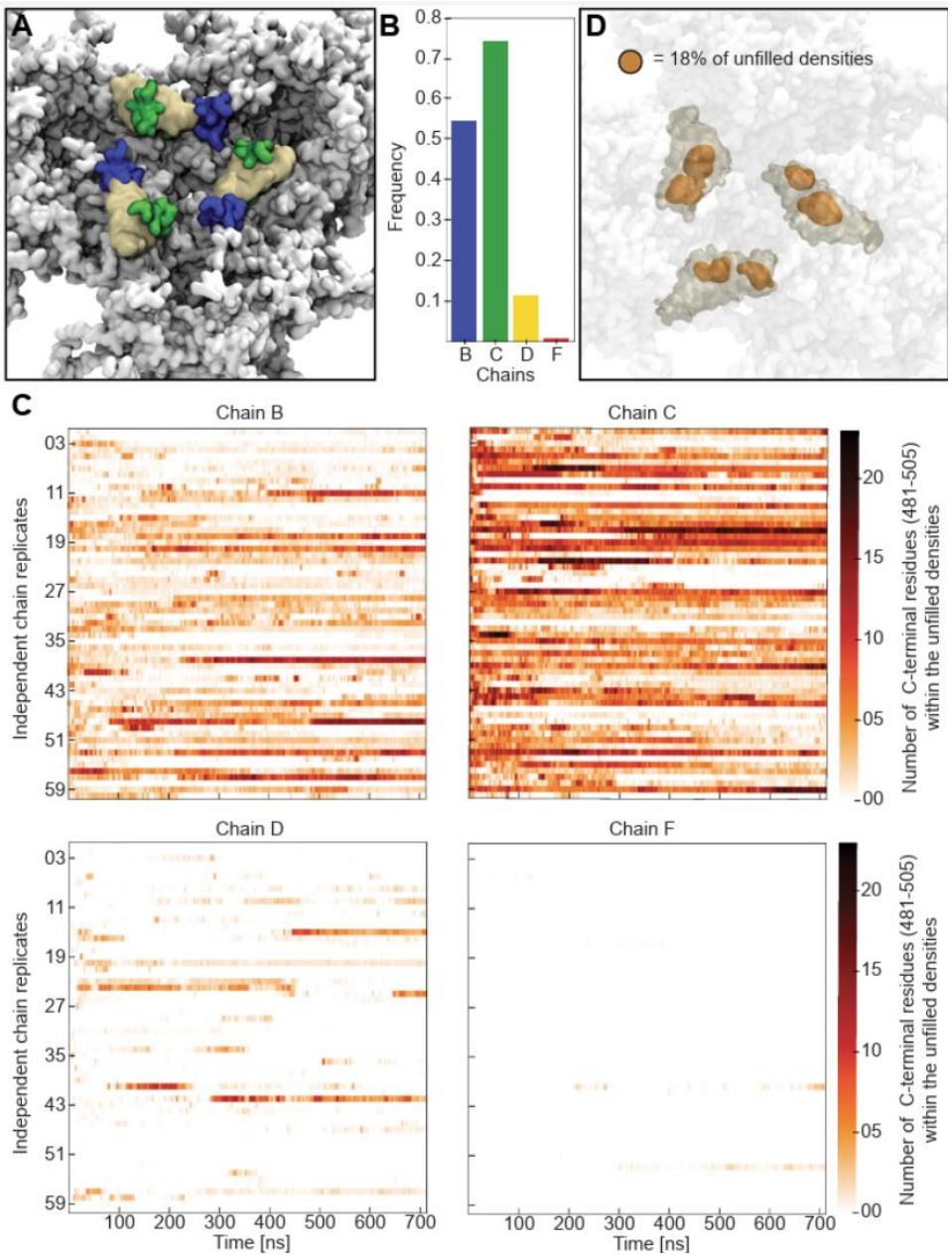
Supplementary Figure 12: According to MD simulations, the N-terminal regions of the C' and F chains tend to form a β-sheet. (A) Representative snapshot of the maximal β-sheet, which spans residues Leu 5 – Leu 13 of chains C' (green) and F (red). Other chains omitted for visual clarity. (B) Occupancy of coil versus β-strand character per residue. Secondary structure assignment used STRIDE. Occupancy data is aggregated over all 60 quasi-equivalent C'–F copies resulting in an effective sampling of 42.6 μs. (C) The fraction of Leu 5 – Leu 13 residue pairs involved in the β-sheet over time. Data for the 60 C'–F copies is sorted by increasing cumulative fraction, where a fraction of 1.0 is defined as all nine residue pairs exhibiting dihedral angles characteristic of a β-strand and hydrogen-bonded pairs sharing at least one such interaction. (D) The average fraction (from panel c) across the 60 C'–F copies, with standard deviation shown in gray. Purple point indicates the β-sheet fraction of the starting model (PDB 7KZF). (E) For each backbone hydrogen bond contributing to the β-sheet (top), the number of C'–F copies simultaneously exhibiting that interaction across the capsid, shown as a function of time (bottom). (F) For each possible auxiliary hydrogen bond that can stabilize the inter-chain interaction, the number of C'–F copies simultaneously exhibiting that interaction across the capsid, shown as a function of time. Source data are provided as a Source Data file.



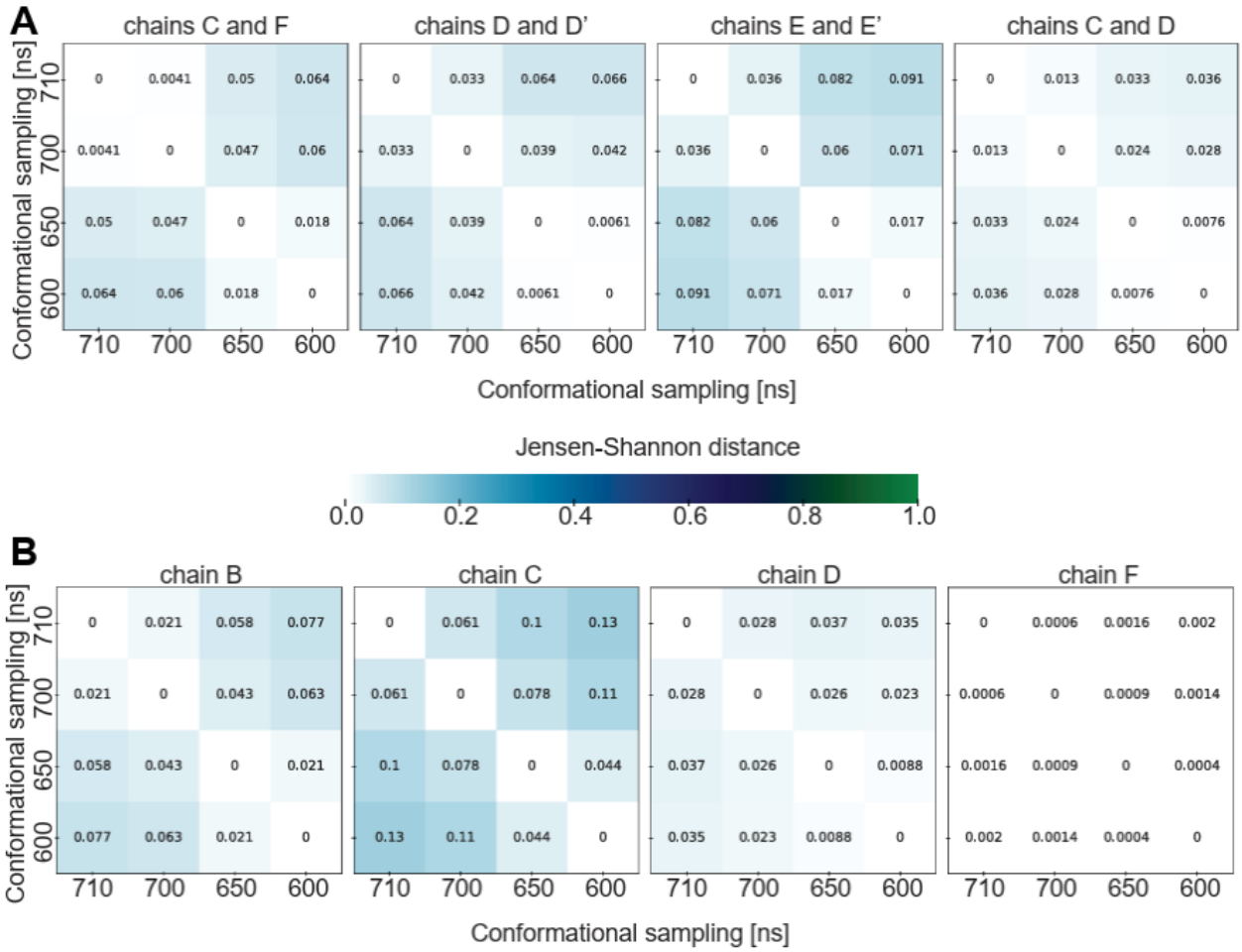
Supplementary Figure 13: Non-L1 density present in a hexavalent capsomer. (A) Inside of hexavalent capsomer with L1 density in blue and unfilled density in gray. Magenta box outlines the volume containing the fishhook density, and the cyan box indicates central density[14]. Both unknown densities have protein-like features. (B and C) Zoomed views of the fishhook and central densities.



Supplementary Figure 14: Amorphous unfilled densities inside the capsid. (A) Center slice of the icosahedral HPV-heparin map at contour 0.25, (B) contour 0.09, and (C) contour 0.07. (D) Zoomed in view of the internal surface of the capsid at contour 0.25, (E) contour 0.09, and (F) contour 0.07. Amorphous unfilled densities were observed on the internal surface of the capsid around the threefold (yellow arrow) and the twofold (orange arrow) axis. The threefold axis of symmetry is indicated with a green triangle.



Supplementary Figure 15: Localization of L1 C-termini during MD simulations. (A) The space sampled by residues 481-503 in MD simulations of the full-length, empty capsid coincides with the location of the unfilled densities along the capsid interior surface (beige), particularly for B (blue) and C (green) chains, for which a representative conformation is shown. (B) Per-chain overlap frequencies of the C-terminal residues with the unfilled experimental density, indicating the relative likelihood of their contribution. Data is aggregated over all 60 independent copies within the capsid for each quasi-equivalent chain. (C) Time series of contacts between residues 481-503 and the unfilled cryoEM densities. (D) Overlapping volume (orange) between unfilled densities and a simulated density calculated from MD conformers of residues 481-503 of chains B and C. Total overlap corresponds to 18% of the volume from unfilled densities. Simulated density was averaged over all trajectory frames for all 20 threefold pores of the capsid, resulting in 14.2 μs of cumulative sampling. Source data are provided as a Source Data file.



Supplementary Figure 16: Pairwise Jensen-Shannon distances (JSD) between distributions comprising different amounts of sampling. (A) Distributions of simultaneous hydrogen bonds between the N-termini of different chains. From left to right: chains C and F, chains D and D', chains E and E', and chains C and D. (B) Distributions of the number of C-terminal residues in contact with unfilled cryoEM densities, for chains B, C, D, and F. JSD values close to zero imply a high degree of similarity between compared distributions. Source data are provided as a Source Data file.

TABLES

Supplementary Table 1: Description of system composition used in MD simulations.

Box shape	Truncated Octahedron
Total atom count	16,818,922
Protein atom count	2,821,620
Water molecule count	4,655,756
Salt concentration	150 mM NaCl